B-083E
C.2.

Statistics Canada Statistique Canada
Ottawa

DESIGN OF THE F.C.C. FARM SURVEY

Susan Ingram
Institutional & Agriculture
Survey Methods Division
March, 1981

DESIGN OF THE F.C.C. FARM SURVEY

1. Introduction

The Farm Credit Corporation (FCC) approached Statistics Canada in 1979 for assistance in designing a survey of Canadian farmers to determine the structure of agricultural assets and debts by province. Specifically, they were interested in comparing FCC borrowers and non-FCC clients to assess if their borrowers had a better balance of credit and assets. As the first stage in the development of a full scale survey, a pilot study was conducted in two FCC field districts, Saskatchewan Region 1 and Quebec Region 6, in February 1980. The sample design of the pilot was formulated by J. Phillips of Institutional and Agriculture Survey Methods Division and is documented in "Design of the FCC Pilot Survey", B. Armstrong, July 1980[1].

One of the ways in which the results of the pilot were analysed involved an evaluation of the pilot survey design. Results of this study are documented in "Report on the Analysis of the FCC Pilot Survey Sample Design", S. Murray, March 1981[2]. Based on the findings outlined in this report, certain changes to the sample design of the full scale survey were recommended.

The FCC decided to undertake a Canada wide Farm Survey during January and February of 1981. The purpose of this report is to document the sample design employed in this survey and to outline the appropriate estimation procedures.

2. Sample Design

The population of interest for this survey consisted of farms which had been given a FCC loan and of active farms on the Central Register of Farms having total sales of $1200 and over. Farms with under $1200 sales were considered to be marginal farmers and were not of interest. It was also decided not to include the two FCC field districts in which the pilot had been undertaken. FCC proposed excluding Quebec Region 6 and Saskatchewan Region 1 and adding on the data from the farms sampled in the pilot survey once they had been updated. Hence, for the purposes of the sample design, farms in Quebec Region 6 and Saskatchewan Region 1 were considered not to be part of the study population.

For each province, FCC files with names and addresses of farmers who had received a FCC loan were matched to active farms on the Central Register. For Quebec and Saskatchewan, farms in the two areas included in the pilot survey were dropped from the files prior to the match. After the match the following three groups were formed within each province: (1) Farms which were on both

the FCC file and the Farm Register; (2) Farms which were on the Farm Register but not on the FCC file; and (3) Farms which were on the FCC file but not on the Central Register. One of the benefits resulting from this grouping was that it ensured that the sample selected would include representation from the subpopulation of farms which had received a FCC loan.

Based on cost and staff constraints the FCC estimated that a sample of approximately 5170 farms could be afforded for the national survey. Furthermore, the sample size allocated to each province was a function of the number of staff which FCC had in the provinces. Given the objective of providing provincial estimates (except for the Atlantic Provinces which were to be considered as one region), some minor modifications were made in the provinces with the smaller sample sizes. This resulted in the sample being allocating among the provinces as indicated in the following table; also included below are the provincial population sizes.

FCC FARM SURVEY

Province	Population Size	Sample Size
British Columbia ...	18,947	250
Alberta	61,763	1,000
Saskatchewan	61,209	950
Manitoba	33,226	550
Ontario	86,877	1,150
Quebec	40,173	980
New Brunswick	5,014	100
Nova Scotia	4,726	77
Prince Edward Island	3,903	100
Newfoundland	688	20
Canada	316,526	5,177

The analysis of the pilot survey results indicated that allocation of the sample proportional to the square root of the population size was a good compromise allocation which should lead to a gain in precision relative to simple random sampling. Therefore, as the first step in the sample allocation, the provincial sample sizes were allocated to each of the 3 groups which had resulted from the match of the FCC and Farm Register files (i.e. matched Farm Register/FCC farms, unmatched Farm Register farms and unmatched FCC farms) proportional to the square root of their population sizes. In comparison to proportional allocation, this form of allocation increases the size of the sample selected from smaller populations, which in this case resulted in a greater number of FCC farms being included in the sample. This was advantageous since a comparison between this group and non-FCC clients was a major objective of the survey.

Farms in groups 1 and 2 (all Central Register farms) were further stratified, if the sample size allocated above permitted, according to an asset value obtained from the census. The total asset value was defined as the sum of the value of land, the value of machinery and the value of livestock. It should be noted that the asset value used for stratification does not correspond to the total asset value obtained in the survey. The latter also includes such items as cash, bonds, savings, accounts receivable, crops for sale, etc. The optimum strata were determined using a technique developed by V.K. Sethi[3]. The number of strata employed was that number which yielded a sample size of at least 10 for all strata, when the sample was allocated proportional to the square root of the stratum population size. In addition, there was a separate stratum for those farms for which census asset value was not available. For farms in group 3 (farms on the FCC file but not on the Farm Register), no information was available for further stratification; hence, this group was sampled from directly.

Further details on the sample design specifications for each province can be found in the appendix; however, the exact asset ranges of the strata cannot be provided due to rules of confidentiality of the census.

The sample was selected using systematic sampling. For Central Register farms (groups 1 and 2) the files from which the samples were selected were ordered geographically (by county/municipality) and within each county/municipality by asset value to ensure that a good spread of the sample throughout the province and by size of farm was achieved. In the unmatched FCC strata, the files were ordered by FCC field/district.

3. Estimation Procedures

The FCC Farm Survey was designed to determine the structure of agricultural assets and debts. The FCC wished to compare their borrowers with non-FCC clients and to make these comparisons for various regions of the country. In order to do this, estimates of various subpopulations are required. These subpopulations are referred to as domains of the population. The estimation procedures given in this section provide the means for obtaining estimates of domain totals and domain means (or proportions).

The procedure consists of applying a weight to each record which represents the number of units in the population which that particular unit is intended to represent. Once the weights have been applied, estimates of domain totals and means for a particular characteristic can be derived simply by summing over all farms in the domain, the product of its weight and the value of the characteristic. Counts of domain units possessing a particular attribute can be obtained by summing over the weights of all farms with that attribute. The mathematical expressions for this procedure as well

as for estimating the coefficient of variation (the estimated standard deviation expressed as a percentage of the estimated value of the characteristic) follow. Standard formulae for estimating domain characteristics for stratified random sampling have been used; however, since the simple random sample variance formula is used to approximate systematic sample variance, the variances may be slightly overestimated.

Notation

h: subscript denoting the h^{th} stratum ($h=1, \ldots, L$)

N_h: total number of units in stratum h

N_d: total number of units in domain d

N_{dh}: total number of units in stratum h in domain d

n_h: number of sample units selected from stratum h

n_{dh}: number of sample units in stratum h in domain d

y_{dhi}: value of the item for the i^{th} sample unit in stratum h in domain d

$\bar{y}_{dh}$: sample mean in stratum h for domain d

$\hat{Y}_d$: estimated domain total

$\bar{Y}_d$: estimated domain mean (or proportion)

$v(\hat{Y}_d)$: estimated variance of estimated domain total

$v(\hat{\bar{Y}}_d)$: estimated variance of estimated domain mean

The weight to be applied to each farm in stratum h is given by $W_h = N_h/n_h$. To obtain estimates of domain totals the following formula is used.

$$\hat{Y}_d = \sum_h^L W_h \sum_i^{n_{dh}} y_{dhi}$$

The estimated variance of the estimated domain total is then

$$v(\hat{Y}_d) = \sum_h^L \frac{N_h (N_h - n_h)}{n_h} s_{dh}^2$$

where $s^2_{dh} = \frac{1}{n_h - 1} \left[\sum_{i}^{n_{dh}} y^2_{dhi} - \frac{\left(\sum_{i}^{n_{dh}} y_{dhi} \right)^2}{n_h} \right]$

and the estimated coefficient of variation is given by

$$c.v.(\hat{Y}_d) = \sqrt{v(\hat{Y}_d)} \; / \; \hat{Y}_d$$

When estimates of domain averages (or proportions) are required, then the following formula is used.

$$\hat{\bar{Y}}_d = \frac{\hat{Y}_d}{\hat{N}_d} = \frac{\sum_{h}^{L} W_h \sum_{i}^{n_{dh}} y_{dhi}}{\sum_{h}^{L} W_h n_{dh}}$$

The estimated variance of $\hat{\bar{Y}}_d$ is given by

$$v(\hat{\bar{Y}}_d) \doteq \frac{1}{\hat{N}_d^2} \sum_{h}^{L} \frac{N_h (N_h - n_h)}{n_h} s'^2_{dh}$$

$$+ \frac{1}{\hat{N}_d^2} \sum_{h}^{L} \frac{N_h (N_h - n_h)}{n_h (n_h - 1)} n_{dh} \left(1 - \frac{n_{dh}}{n_h}\right) \left(\bar{y}_{dh} - \hat{\bar{Y}}_d\right)^2$$

where $\hat{N}_d = \sum_{h}^{L} W_h \, n_{dh}$

$$s'^2_{dh} = \frac{1}{n_h - 1} \left[\sum_{i}^{n_{dh}} y^2_{dhi} - \frac{\left(\sum_{i}^{n_{dh}} y_{dhi} \right)^2}{n_{dh}} \right]$$

$$\bar{y}_{dh} = \frac{1}{n_{dh}} \sum_{i}^{n_{dh}} y_{dhi}$$

The estimated coefficient of variation is then given by

$$c.v.(\hat{\bar{Y}}_d) = \sqrt{v(\hat{\bar{Y}}_d)} \; / \; \hat{\bar{Y}}_d$$

It should be noted that the second term of the variance expression is small if the domain is almost as large as the complete population and, in fact, disappears when estimating $\bar{Y}$ for the whole population.

The above formulae should be used to produce domain estimates within each province. To obtain estimates of characteristics (and estimated variances) for larger regions or for Canada as a whole, the provincial estimates should be summed. As well, the estimates for Quebec Region 6 and Saskatchewan Region 1 should be determined by applying the formulae to the updated pilot survey data. The resulting estimates should then be added to the appropriate estimate from this survey to produce the estimates for Quebec and Saskatchewan, respectively.

During the course of the field work, it was discovered that there were a few duplicate operations included in the sample. To adjust for such duplication in the estimation procedures, one should replace y_{dhi} with y_{dhi}/a_{hi} in the estimation formulae provided above (where a_{hi} is the number of times the operation is included in the sample). If the population values A_{hi} (the number of times the operation appears on the frame) were known, the estimator would be unbiased. However, since the number of times the unit is included in the sample is less than or equal to the true number of times the unit is duplicated in the frame, the estimator which uses a_{hi} is biased in the positive direction. In practice this bias is expected to be small if efforts are made to unduplicate the frame since the number of duplicates on the frame will probably be minimal with respect to the size of the frame. Although it is possible to derive an unbiased estimator when the A_{hi} are unknown, the resulting estimator is complex and its mean squared error is greater than that of the estimator employing a_{hi}[4]. Therefore, since the files from which the sample for the FCC survey were selected had been unduplicated using name and address prior to the sample selection, the estimator which replaces y_{dhi} with y_{dhi}/a_{hi} should be employed to adjust for the 3-4 duplicate operations discovered during the course of the survey.

APPENDIX

FCC FARM SURVEY SAMPLE DESIGN SPECIFICATIONS

SASKATCHEWAN

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	5,626	182
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	175	13
		3	3,072	56
		4	2,770	53
		5	2,371	49
		6	1,651	41
		7	933	31
		8	175	13
	Farms not on FCC file	9	3,063	55
		10	13,140	114
		11	11,378	106
		12	8,229	91
		13	5,780	76
		14	2,422	49
		15	424	21
TOTAL			61,209	950

MANITOBA

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	3,231	112
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	79	8
		3	1,219	30
		4	1,163	30
		5	799	24
		6	417	18
		7	155	11
	Farms not on FCC file	8	2,181	41
		9	9,635	87
		10	7,662	78
		11	4,771	61
		12	1,655	36
		13	259	14
TOTAL			33,226	550

ONTARIO

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	7,366	221
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	234	16
		3	2,488	52
		4	2,325	50
		5	1,824	45
		6	1,446	40
		7	759	29
		8	229	16
	Farms not on FCC file	9	9,029	101
		10	20,117	151
		11	18,604	145
		12	12,582	120
		13	6,992	89
		14	2,474	53
		15	408	22
TOTAL			86,877	1,150

QUEBEC

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	2,274	158
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	152	18
		3	1,087	48
		4	1,237	51
		5	863	43
		6	483	32
		7	126	17
	Farms not on FCC file	8	3,050	75
		9	8,641	127
		10	8,706	127
		11	6,973	114
		12	4,445	91
		13	1,951	60
		14	185	19
TOTAL			40,173	980

NEW BRUNSWICK

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	364	18
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	373	19
	Farms not on FCC file	3	1,566	23
		4	2,330	28
		5	381	12
TOTAL			5,014	100

NOVA SCOTIA

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	99	10
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	97	10
	Farms not on FCC file	3	1,388	19
		4	2,568	26
		5	574	12
TOTAL			4,726	77

PRINCE EDWARD ISLAND

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	396	21
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	421	21
	Farms not on FCC file	3	571	16
		4	2,184	30
		5	331	12
TOTAL			3,903	100

NEWFOUNDLAND*

Stratum Description	Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on Central Register which did not match to a farm on the FCC file	1	613	10
FCC farms	2	75	10
TOTAL		688	20

*Farms in group 1 (Farms on both files) and group 3 (Farms on FCC file only) were grouped to form stratum 2 due to small population and sample sizes.

BRITISH COLUMBIA

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	1,607	51
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	670	33
	Farms not on FCC file	3	5,648	53
		4	7,859	62
		5	2,842	38
		6	321	13
TOTAL			18,947	250

ALBERTA

Stratum Description		Stratum I.D.	Population Size (N_h)	Sample Size (n_h)
All farms on FCC file which did not match to a farm on the Central Register		1	5,474	191
All farms on Central Register with '76 Census sales ≥$1200	Farms on FCC file	2	123	12
		3	2,723	55
		4	2,483	53
		5	1,926	47
		6	1,251	38
		7	728	29
		8	258	17
	Farms not on FCC file	9	2,794	57
		10	16,621	140
		11	12,281	121
		12	8,592	101
		13	4,372	72
		14	1,774	46
		15	363	21
TOTAL			61,763	1,000

REFERENCES

1) Armstrong, B. "Design of the FCC Pilot Survey", July, 1980

2) Murray, S. "Report on the Analysis of the FCC Pilot Survey Sample Design", March, 1981

3) Sethi, V.K. "A Note on Optimum Stratifications for Estimating the Population Means" Australian Journal of Statistics, 5 (1963)

4) Shimizu, I. "Estimation with Duplicate Sample Units from Stratified Samples", American Statistical Association Proceedings, (1975)